



ANALISIS KEPUASAN PENGGUNA APLIKASI NEOBANK MELALUI ANALISIS SENTIMEN DAN KLASIFIKASI MULTILABEL DENGAN LOGISTIC REGRESSION PADA KERANGKA EUCS

Stevi Aprilianti Cahyani¹⁾, Wiyli Yustanti²⁾

¹⁾ Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: steviaprilianti.22027@mhs.unesa.ac.id

²⁾ Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: wiyliyustanti@unesa.ac.id

Abstract

The rapid development of digital banking services in Indonesia has led to an increased adoption of mobile banking applications as the primary medium for financial transactions. Neobank, as one of the mobile banking applications with a high number of downloads, exhibits a discrepancy between usage intensity and user satisfaction, as reflected in its ratings and user reviews. Various user reviews indicate issues related to technical constraints, system accuracy, ease of use, and service timeliness. This condition highlights the importance of conducting a comprehensive evaluation of user satisfaction. This study aims to analyze user satisfaction with the Neobank application through the integration of sentiment analysis and multilabel classification based on the End-User Computing Satisfaction (EUCS) framework. The dataset consists of 10,392 user reviews collected from the Google Play Store and App Store. The research stages include text preprocessing, sentiment labeling using a pre-trained Indonesian RoBERTa Base Sentiment Classifier, and EUCS dimension labeling using a fuzzy string matching approach based on keywords. Subsequently, a multilabel classification model was developed using Logistic Regression with a One-vs-Rest (OVR) approach and TF-IDF features. The evaluation results on the test data demonstrate a precision of 0.940, a recall of 0.809, and an F1-score of 0.869 (macro average). The k-fold cross-validation results (k=5) indicate stable model performance, with the highest average F1-score achieved in the Ease of Use dimension (0.927). The model was then implemented into a web-based system to perform sentiment prediction, EUCS multilabel classification, and generate improvement recommendations using a Large Language Model (LLM).

Keywords: End User Computer Satisfaction; Large Language Model; Multilabel Classification; Sentiment Analysis; User Satisfaction

Abstrak

Perkembangan layanan perbankan digital di Indonesia mendorong meningkatnya penggunaan aplikasi mobile banking sebagai sarana utama transaksi keuangan. Neobank sebagai salah satu aplikasi mobile banking yang memiliki jumlah unduhan tinggi, namun tingkat rating dan ulasan pengguna menunjukkan adanya ketidaksesuaian antara intensitas penggunaan dan kepuasan pengguna. Berbagai ulasan pengguna mengindikasikan adanya permasalahan terkait kendala teknis, keakuratan sistem, kemudahan penggunaan, serta ketepatan waktu layanan. Kondisi tersebut menunjukkan pentingnya evaluasi kepuasan pengguna secara komprehensif. Penelitian ini bertujuan untuk menganalisis kepuasan pengguna aplikasi Neobank melalui integrasi antara analisis sentimen dan klasifikasi multilabel berdasarkan kerangka End-User Computing Satisfaction (EUCS). Data penelitian berupa 10.392 ulasan pengguna yang diperoleh dari Google Play Store dan App Store. Tahapan penelitian meliputi preprocessing teks, pelabelan sentimen menggunakan model pre-trained Indonesian RoBERTa Base Sentiment Classifier, serta pelabelan dimensi EUCS menggunakan pendekatan fuzzy string matching berbasis keyword. Selanjutnya, dibangun model klasifikasi multilabel menggunakan Logistic Regression dengan pendekatan One-vs-Rest (OVR) dan fitur TF-IDF. Hasil evaluasi pada data uji menunjukkan precision sebesar 0,940, recall sebesar 0,809, dan F1-score sebesar 0,869 (macro average). Hasil k-fold cross validation (k=5) menunjukkan performa model yang stabil dengan rata-rata F1-score tertinggi pada dimensi Ease of Use sebesar 0,927. Model kemudian diimplementasikan ke dalam sistem berbasis web untuk melakukan prediksi sentimen, klasifikasi multilabel EUCS, serta menghasilkan rekomendasi perbaikan berbasis Large Language Model (LLM).

Kata Kunci: Analisis Sentimen; End User Computer Satisfaction; Kepuasan Pengguna, Large Language Model; Multilabel Classification.



PENDAHULUAN

Pesatnya perkembangan teknologi informasi telah mendorong transformasi digital di berbagai sektor, salah satunya adalah perbankan, di mana layanan *mobile banking* kini menjadi kebutuhan mendasar bagi masyarakat dalam melakukan transaksi layanan keuangan secara lebih cepat, mudah, dan efisien. (Johri 2023) Namun, fenomena yang terjadi pada aplikasi Neobank menunjukkan adanya kesenjangan antara tingginya jumlah unduhan dengan tingkat kepuasan pengguna yang masih relatif rendah. Berbagai ulasan pengguna di *platform* digital seperti *Google Play Store* dan *App Store* secara konsisten menyoroti kendala teknis, seperti kesulitan *login*, keterlambatan transaksi, notifikasi yang tidak akurat, hingga rating rendah yang mengindikasikan bahwa kualitas layanan aplikasi tersebut belum sepenuhnya memenuhi ekspektasi pengguna. Oleh karena itu, diperlukan penelitian mendalam untuk mengevaluasi tingkat kepuasan pengguna secara komprehensif untuk memahami akar permasalahan dari aplikasi tersebut.

Salah satu pendekatan yang umum digunakan adalah *End-User Computing Satisfaction (EUCS)*. EUCS mengukur kepuasan melalui lima dimensi, yaitu *Ease of Use*, *Content*, *Accuracy*, *Timeliness*, dan *Format* (Doll and Torkzadeh 1988). Namun penelitian yang menggunakan pendekatan EUCS kebanyakan masih menggunakan metode pengumpulan data yang berupa survei kuesioner daring, sehingga memiliki keterbatasan signifikan, terutama dalam verifikasi populasi target yang sebenarnya, serta munculnya bias *self-selection* yaitu ketika responden yang berpartisipasi dipilih secara sukarela sehingga tidak mencerminkan seluruh populasi dan kurang mampu menangkap opini pengguna yang bersifat spontan, kontekstual, dan *real-time* (Kennedy, Jensen, and Jensen 2022; Rubenstein and Furnier 2021).

Sebagai alternatif survei, ulasan pengguna memberikan gambaran pengalaman yang spontan dan kontekstual secara langsung (Sukmana et al. 2024). Dalam memproses data teks berukuran besar, analisis sentimen menjadi metode utama yang digunakan. Model *IndoRoBERTa-Base-Sentiment-Classifer* dipilih sebagai metode utama untuk mengolah ulasan pengguna Neobank karena terbukti memiliki akurasi klasifikasi yang lebih unggul dibandingkan model lain seperti XLM-RoBERTa, DistilBERT, maupun SVM (Anindya and Kaesmetan 2025; Djaelani and Thamrin 2025), sehingga mampu memberikan gambaran pengalaman pengguna yang lebih spontan dan kontekstual sebelum dikategorikan ke dalam dimensi EUCS secara multilabel dengan menggunakan Logistic Regression karena memiliki performa yang stabil, interpretabilitas yang baik, efisiensi komputasi dan memberikan akurasi lebih tinggi dibandingkan *Random Forest* dan *Support Vector Classification (SVC)*. (Ibadurrahman et al. 2025; Pratama et al. 2022; Sonjaya et al. 2022; Syatriani, Dinata, and ArRazi 2025)

Di sisi lain, pemanfaatan *Large Language Model (LLM)* dalam pengolahan bahasa alami terbukti unggul dalam menangani tugas generatif, interpretatif, dan kontekstual pada skenario dinamis dibandingkan pendekatan *rule-based* tradisional. Sebagaimana

ditunjukkan melalui efektivitas integrasi Groq AI pada sistem evaluasi kinerja, pengembangan *chatbot* akademik, optimalisasi sistem asisten berbasis suara, serta penggunaan *Case-Based Reasoning (CBR)* untuk rekomendasi berbasis konteks. (Alfarisi et al. 2026; Alvin, Robet, and Tarigan 2025; Kusumastuti, Oktafiani, and Putra 2025; Supriana et al. 2025). Mengingat belum banyaknya penelitian yang mengintegrasikan LLM dengan model klasifikasi teks untuk mendukung evaluasi kepuasan pengguna berbasis kerangka EUCS, penelitian ini memanfaatkan LLM melalui integrasi Groq AI untuk menghasilkan rekomendasi perbaikan secara otomatis berdasarkan hasil prediksi sentimen dan dimensi EUCS, sehingga sistem tidak hanya mampu melakukan klasifikasi tetapi juga menyediakan rekomendasi tindakan yang aplikatif bagi pengembang.

Berdasarkan uraian di atas, penelitian mengenai “Analisis Kepuasan Pengguna Aplikasi NeoBank melalui Analisis Sentimen dan Klasifikasi *Multilabel* dengan *Logistic Regression* pada Kerangka EUCS” sangat penting untuk dilakukan. Penelitian ini diharapkan tidak hanya memberikan kontribusi di bidang akademis dalam pengembangan kajian kepuasan pengguna dan pembuatan model dengan mengintegrasikan antara analisis sentimen dan kerangka EUCS, tetapi juga dapat memberikan manfaat yang praktis bagi pengembang aplikasi dalam meningkatkan kualitas layanan *digital banking*. Dengan pendekatan analisis yang komprehensif, hasil penelitian ini dapat memberikan gambaran yang lebih akurat tentang persepsi pengguna sekaligus menjadi masukan strategis bagi pengembangan layanan Neobank agar lebih responsif, efisien, dan sesuai dengan kebutuhan pengguna.

TINJAUAN PUSTAKA

Untuk memberikan landasan pemahaman yang komprehensif dalam penelitian ini, bagian berikut akan menguraikan beberapa teori dan konsep utama yang menjadi acuan. Pembahasan ini mencakup End User Computing Satisfaction sebagai kerangka evaluasi, teknik analisis data melalui Klasifikasi Teks dan Analisis Sentimen, serta Integrasi LLM dalam Sistem Rekomendasi sebagai kerangka pengembangan solusi.

End User Computing Satisfaction

End-User Computing Satisfaction (EUCS) yang dikembangkan oleh (Doll and Torkzadeh 1988) merupakan kerangka teoretis utama untuk mengevaluasi kepuasan pengguna terhadap implementasi sistem informasi melalui dimensi *Content*, *Accuracy*, *Format*, *Ease of Use*, dan *Timeliness*. Kerangka EUCS telah banyak diterapkan dalam evaluasi sistem informasi dan aplikasi digital seperti pada penelitian oleh (Prabawanti and Sihombing 2023), (Ernawati, Faroqi, and Aulia 2025), (Umar, Fitriansyah, and Ar-Rasyid 2023), (Aini et al. 2023), serta (Ramma and Nuryana 2025). Namun, kebanyakan masih menggunakan metode pengumpulan data yang berupa survei kuesioner daring. Metode ini memiliki beberapa keterbatasan signifikan, terutama dalam verifikasi populasi target yang sebenarnya, serta munculnya bias *self-selection*, yaitu ketika responden yang berpartisipasi dipilih secara sukarela sehingga tidak mencerminkan seluruh populasi (Rubenstein



and Furnier 2021). Selain itu, penelitian metodologis juga menegaskan bahwa survei daring seringkali menghadapi tantangan dalam validitas pengukuran dan keterwakilan data (Kennedy et al. 2022). Kondisi ini membuat survei kurang mampu menangkap opini pengguna yang bersifat spontan, kontekstual, dan *real-time*. Sejalan dengan temuan penelitian yang menekankan pentingnya umpan balik alami, penelitian ini beralih menggunakan data tidak terstruktur berupa ulasan pengguna (*user-generated content*) dari platform digital. Data ini dinilai lebih autentik karena mencerminkan pengalaman pengguna yang spontan, kontekstual, dan tidak terkontrol oleh instrumen kuesioner, sehingga mampu memberikan *insight* yang lebih kaya mengenai permasalahan teknis dan fungsional aplikasi Neobank secara nyata.

Klasifikasi Teks dan Analisis Sentimen

Analisis Sentimen merupakan cabang dari penelitian text mining yang melakukan proses pengklasifikasian dokumen teks. Analisis sentimen dapat melakukan ekstraksi pendapat, emosi, dan evaluasi tertulis seseorang tentang topik tertentu menggunakan teknik pemrosesan Bahasa alami (Idris I, Mustofa Y, and Salihi I 2023). Salah satu metode yang bisa digunakan yaitu RoBERTa, merupakan salah satu model NLP yang dilatih dalam korpus bahasa tertentu, kompleks dan mampu menghasilkan representasi teks dengan baik, pada penelitian (Sihombing et al. 2024), (Anindya and Kaesmetan 2025) dan Djaelani & Thamrin, (2025) menunjukkan bahwa *Indonesian-RoBERTa-Base-Sentiment-Classifier* memiliki kinerja lebih unggul dibandingkan dengan model SVM, *XLM-RoBERTa* dan *DistilBERT*. Untuk klasifikasi multilabel, penelitian ini menerapkan *Logistic Regression* dengan pendekatan *One-vs-Rest* (OvR). Pemilihan *Logistic Regression* didasarkan pada karakteristiknya yang memiliki efisiensi komputasi tinggi, tingkat interpretabilitas yang baik, serta performa stabil dalam menangani ketidakseimbangan kelas pada dataset ulasan aplikasi, sebagaimana dibuktikan oleh penelitian (Sonjaya et al. 2022) menegaskan bahwa *Logistic Regression* tetap menunjukkan performa yang seimbang pada dataset dengan ketidakseimbangan kelas, terutama ketika dikombinasikan dengan teknik oversampling seperti SMOTE. Temuan tersebut diperkuat oleh (Pratama et al. 2022) yang menunjukkan bahwa *Logistic Regression* memiliki kemampuan generalisasi yang baik serta metrik evaluasi yang kompetitif dalam konteks sistem rekomendasi.

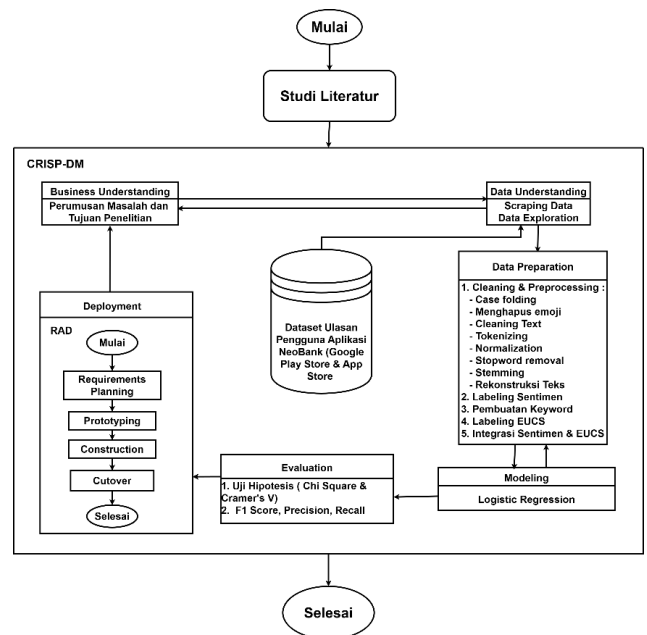
Integrasi LLM dalam Sistem Rekomendasi

Kemajuan dalam *Natural Language Processing* saat ini telah membawa *Large Language Model* (LLM) ke garis depan sistem interaktif. Penelitian oleh Kusumastuti et al. (2025) dan Supriana et al. (2025) membuktikan bahwa integrasi LLM mampu menghasilkan respons yang dinamis, kontekstual, dan solutif dibandingkan dengan sistem berbasis aturan tradisional (*rule-based system*). Meskipun banyak riset berfokus pada *chatbot* umum, masih terdapat celah penelitian dalam mengoperasikan model generatif untuk menyusun rekomendasi perbaikan spesifik berdasarkan klasifikasi dimensi EUCS. Oleh karena itu,

penelitian ini mengisi *research gap* tersebut dengan mengintegrasikan model klasifikasi *Logistic Regression* untuk mendeteksi permasalahan dimensi EUCS, yang kemudian diproses oleh LLM untuk *generate* rekomendasi tindakan perbaikan yang otomatis dan relevan bagi pengembang aplikasi.

METODE

Penelitian ini menggabungkan dua kerangka kerja, yaitu CRISP-DM sebagai framework utama dalam proses analisis data dan RAD sebagai pendekatan pengembangan website deteksi ulasan, dimana tahapannya adalah sebagai berikut:



Gambar 1. Framework Penelitian

Business Understanding

Neobank telah digunakan oleh jutaan pengguna, namun masih terdapat banyak ulasan yang menunjukkan berbagai keluhan terkait kualitas layanan, seperti kendala login, keterlambatan transaksi, kesalahan notifikasi, serta masalah pada fitur aplikasi. Selama ini, evaluasi kepuasan pengguna umumnya dilakukan melalui survei atau kuesioner yang memiliki keterbatasan dari segi jumlah responden, waktu pengumpulan data, dan potensi bias. Di sisi lain, ulasan pengguna pada Google Play Store dan App Store merupakan sumber informasi yang bersifat spontan, autentik, dan terus bertambah setiap waktu sehingga dapat dimanfaatkan sebagai dasar evaluasi kualitas layanan. Oleh karena itu, penelitian ini bertujuan membangun model yang mampu menganalisis sentimen ulasan pengguna, mengklasifikasikan komentar ke dalam dimensi *End-User Computing Satisfaction* (EUCS), serta menghasilkan rekomendasi perbaikan secara otomatis. Hasil penelitian diharapkan dapat membantu pengembang Neobank dalam memahami kebutuhan pengguna dan menentukan prioritas peningkatan kualitas layanan berdasarkan data ulasan pengguna.



Data Understanding

Tahapan penelitian ini dimulai dengan pengumpulan data ulasan pengguna aplikasi NeoBank melalui proses *scraping* dari Google Play Store dan App Store dengan batasan maksimal 15.000 ulasan yang relevan hingga 10 Oktober 2025. Proses *scraping* ini mencakup pengambilan atribut penting seperti identitas unik, nama pengguna, teks ulasan, *rating*, dan tanggal unggahan yang kemudian disaring dari data duplikat serta digabungkan ke dalam format csv. Selanjutnya, dilakukan eksplorasi data untuk memastikan kualitas dan reliabilitas dataset melalui pemeriksaan struktur atribut, analisis sebaran *rating* dan panjang teks guna memahami pola opini, serta pembersihan data dari nilai kosong maupun entri duplikat guna memastikan dataset siap untuk dianalisis lebih lanjut

Data Preparation

Tahap *Data Preparation* berfokus pada proses pembersihan dan penyiapan data agar siap digunakan dalam tahap analisis dan pemodelan. Tujuan utama dari tahap ini adalah memastikan bahwa data yang digunakan memiliki kualitas baik, bebas dari kesalahan, serta dalam format yang sesuai untuk diproses ketika melakukan modelling nanti. Pada penelitian ini terdapat lima Langkah dalam persiapan data yaitu i) Preprocessing Data, ii) Labeling Sentimen, iii) Pembuatan Keyword EUCS, iv) Labeling EUCS, v) Integrasi Sentimen dan EUCS.

Preprocessing Data dilakukan untuk membersihkan teks melalui proses *case folding*, *cleaning text*, *tokenizing*, *normalization*, *stopword removal*, *stemming*, dan *rekonstruksi teks*. Selanjutnya, Labeling Sentimen dilakukan menggunakan model *Indonesian RoBERTa Base Sentiment Classifier* untuk mengategorikan ulasan ke dalam sentimen positif, netral, atau negatif.

Selanjutnya, dilakukan pembuatan *keyword* EUCS berdasarkan studi literatur dan observasi ulasan berdasarkan lima dimensi EUCS. Labeling EUCS kemudian diterapkan dengan metode *Fuzzy String Matching* dengan menggunakan algoritma *Levenshtein Distance* untuk mengidentifikasi dimensi yang dibahas dalam ulasan dengan *threshold* 75%. Untuk hasil labeling dapat dilihat pada Tabel 1.

Tabel 1. Contoh Hasil Labeling Dimensi EUCS

Ulasan	Ease of Use	Content	Accuracy	Timeliness	Format	Hasil
Tampilan aplikasinya sangat bagus dan aplikasi mudah digunakan	1	0	0	0	1	Ease of Use, Format

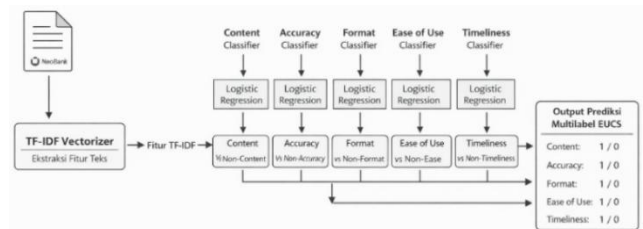
Terakhir, Integrasi Sentimen dan EUCS dilakukan dengan mengonversi label sentimen menjadi skor numerik (1=negatif, 2=netral, 3=positif) dan menggabungkannya dengan label dimensi EUCS. Hasil integrasi ini membentuk dataset terstruktur yang merepresentasikan skor kepuasan pada setiap dimensi EUCS, menyerupai data kuisioner skala Likert yang dapat dilihat pada Tabel 2.

Tabel 2. Contoh Hasil Integrasi Sentimen dan Dimensi EUCS

Ulasan	Ease of Use	Content	Accuracy	Timeliness	Format	Satisfaction
Tampilan aplikasinya sangat bagus dan aplikasi mudah digunakan	3	2	2	2	3	Positif

Modeling

Tahap *Modeling* dalam penelitian ini berfokus pada pembangunan model klasifikasi *multilabel* menggunakan pendekatan *One-vs-Rest* dengan algoritma *Logistic Regression* untuk mengidentifikasi dimensi EUCS pada ulasan pengguna aplikasi Neobank, di mana setiap dimensi diperlakukan sebagai target klasifikasi biner yang independen. Untuk memastikan reliabilitas dan mengurangi bias evaluasi, model divalidasi menggunakan teknik *5-Fold Cross Validation* yang membagi dataset ke dalam lima subset untuk pelatihan dan pengujian secara bergantian, serta seluruh *pipeline* model yang mencakup proses ekstraksi fitur *TF-IDF* dan *classifier* kemudian disimpan dalam format *pickle* (.pkl) menggunakan pustaka *joblib* agar dapat digunakan kembali pada tahap *deployment*. Untuk memperjelas alur kerja pendekatan *One-vs-Rest*, berikut merupakan visualisasi mekanisme klasifikasi *multilabel* disajikan pada Gambar 2.



Gambar 2. Alur Kerja Logistic Regression OvR

Evaluation

Tahap *Evaluation* dalam penelitian ini dilakukan untuk menilai kinerja model *Logistic Regression* dalam memprediksi lima dimensi EUCS menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk menyeimbangkan presisi dan kemampuan deteksi label positif. Selain evaluasi performa model, dilakukan Uji Chi-Square untuk menganalisis signifikansi hubungan antara masing-masing dimensi EUCS dengan tingkat kepuasan pengguna. Uji ini dilakukan dengan membandingkan frekuensi yang diamati di lapangan terhadap frekuensi yang diharapkan secara teoretis menggunakan rumus:

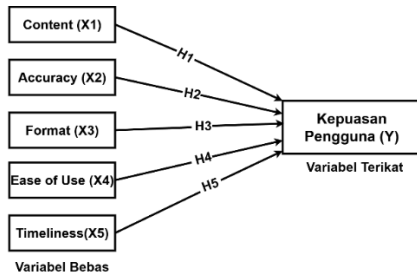
$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

Selain itu juga dilakukan uji Cramer's V untuk mengukur kekuatan hubungan antara dua variabel setelah diketahui adanya hubungan signifikan dari hasil uji Chi-Square. Rumus Cramér's V adalah sebagai berikut:



$$V = \sqrt{\frac{x^2}{N \times (k - 1)}} \quad (2)$$

Untuk setiap dimensi EUCS, hipotesis yang diuji dirumuskan sebagai berikut:



Gambar 3. Model Konseptual Uji Hipotesis

H_0 : Tidak terdapat hubungan yang signifikan antara dimensi EUCS dengan tingkat kepuasan pengguna aplikasi Neobank.

H_1 : Terdapat hubungan yang signifikan antara dimensi EUCS dengan tingkat kepuasan pengguna aplikasi Neobank.

Tabel 3. Rumusan Hipotesis

Dimensi	Hipotesis Nol (H_0)	Hipotesis Alternatif (H_1)
Content (X_1)	Tidak terdapat hubungan yang signifikan antara dimensi <i>Content</i> dengan tingkat kepuasan pengguna.	Terdapat hubungan yang signifikan antara dimensi <i>Content</i> dengan tingkat kepuasan pengguna.
Accuracy (X_2)	Tidak terdapat hubungan yang signifikan antara dimensi <i>Accuracy</i> dengan tingkat kepuasan pengguna.	Terdapat hubungan yang signifikan antara dimensi <i>Accuracy</i> dengan tingkat kepuasan pengguna.
Format (X_3)	Tidak terdapat hubungan yang signifikan antara dimensi <i>Format</i> dengan tingkat kepuasan pengguna.	Terdapat hubungan yang signifikan antara dimensi <i>Format</i> dengan tingkat kepuasan pengguna.
Ease of Use (X_4)	Tidak terdapat hubungan yang signifikan antara dimensi <i>Ease of Use</i> dengan tingkat kepuasan pengguna.	Terdapat hubungan yang signifikan antara dimensi <i>Ease of Use</i> dengan tingkat kepuasan pengguna.
Timeliness (X_5)	Tidak terdapat hubungan yang signifikan antara dimensi <i>Timeliness</i> dengan tingkat kepuasan pengguna.	Terdapat hubungan yang signifikan antara dimensi <i>Timeliness</i> dengan tingkat kepuasan pengguna.

Deployment

Tahap *deployment* dalam penelitian ini bertujuan untuk mengimplementasikan model analisis sentimen dan klasifikasi multilabel dimensi EUCS ke dalam antarmuka sistem berbasis *website* dengan mengadopsi pendekatan *Rapid Application Development* (RAD) untuk memastikan proses pengembangan yang iteratif, terstruktur, dan efisien. Sistem ini dirancang menggunakan arsitektur *client-server* yang mengintegrasikan *framework* Flask pada *backend* untuk menangani pemrosesan data meliputi *preprocessing* teks, analisis sentimen menggunakan IndoRoBERTa, prediksi dimensi EUCS melalui *Logistic Regression*, serta rekomendasi perbaikan kontekstual berbasis *Large Language Model* (LLM) menggunakan Groq API dengan komponen *frontend* berbasis HTML, CSS, dan JavaScript

sebagai media interaksi pengguna. Melalui pengujian *Black Box*, sistem telah divalidasi mampu menjalankan fungsionalitas utama secara optimal, baik untuk mekanisme input *single comment* maupun *batch comment*, sehingga menghasilkan keluaran berupa visualisasi hasil analisis yang informatif dan mendukung pengambilan keputusan untuk evaluasi kualitas aplikasi.

HASIL DAN PEMBAHASAN

Persepsi Kepuasan Pengguna Aplikasi Neobank

Tabel 4. Persepsi Kepuasan Pengguna

Dimensi EUCS	Negatif	Netral	Positif
Content	5332	6935	3042
Accuracy	5246	6689	2964
Format	4984	6345	2891
Ease of Use	5335	6894	3043
Timeliness	5258	6709	2954

Pada Tabel 4 menunjukkan bahwa seluruh dimensi EUCS pada aplikasi Neobank didominasi oleh sentimen netral, yang mengindikasikan bahwa sebagian besar ulasan pengguna bersifat informatif atau deskriptif tanpa ekspresi kepuasan yang eksplisit. Pada dimensi *Content* tercatat total 15.309 ulasan, sedangkan dimensi *Ease of Use* mencatatkan 15.272 ulasan, yang menunjukkan bahwa kedua aspek tersebut paling banyak dibahas oleh pengguna. Namun, tingginya jumlah ulasan ini tetap disertai dengan munculnya sentimen negatif yang signifikan pada aspek kualitas informasi dan pengalaman penggunaan aplikasi. Sementara itu, dimensi *Accuracy*, *Format*, dan *Timeliness* memiliki jumlah ulasan yang lebih seimbang, namun tetap mengandung sentimen negatif yang menandakan bahwa aspek akurasi informasi, tampilan antarmuka, serta ketepatan waktu layanan masih belum memenuhi harapan pengguna. Secara keseluruhan, belum dominannya sentimen positif pada seluruh dimensi EUCS menandakan bahwa aplikasi Neobank masih memerlukan peningkatan pada berbagai aspek sistem untuk meningkatkan kepuasan pengguna.

Tabel 5. Hasil Uji Hipotesis

Dimensi	Chi-Square	p-value	Cramer's V	Hasil
Content	90.064	0.00	0.077	Tolak H_0
Accuracy	347.389	0.00	0.150	Tolak H_0
Format	36.774	0.00	0.049	Tolak H_0
Ease of Use	68.029	0.00	0.067	Tolak H_0
Timeliness	125.851	0.00	0.090	Tolak H_0

Sementara untuk hasil uji hipotesis menunjukkan bahwa seluruh dimensi EUCS memiliki hubungan yang signifikan dengan kepuasan pengguna aplikasi, yang dibuktikan dengan nilai *p-value* sebesar 0,000 ($< 0,05$) pada seluruh dimensi. Meskipun hubungan tersebut signifikan secara statistik, nilai Cramér's V menunjukkan bahwa kekuatan hubungan berada pada kategori lemah hingga sedang. Dimensi *Accuracy* memiliki pengaruh paling dominan dengan nilai Cramér's V sebesar 0,150, diikuti oleh *Timeliness* (0,090), *Content* (0,077), *Ease of Use*



(0,067), dan *Format* dengan nilai terendah sebesar 0,049. Secara keseluruhan, dimensi yang berkaitan langsung dengan fungsi utama aplikasi, khususnya *Accuracy* dan *Timeliness*, terbukti memiliki pengaruh yang relatif lebih besar terhadap kepuasan pengguna dibandingkan aspek lainnya.

Model Optimal Prediksi Multilabel

Dalam pengembangan model prediksi multilabel dimensi *End-User Computing Satisfaction* (EUCS), dataset dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% untuk memastikan generalisasi model yang optimal. Berdasarkan hasil evaluasi yang dapat dilihat pada Tabel 6, model mencapai *precision* sebesar 0,940, *recall* sebesar 0,809, dan *F1-score* sebesar 0,869 pada data uji, dengan penurunan yang wajar dari data latih sehingga model dinyatakan bebas dari *overfitting*.

Tabel 6. Hasil Training dan Testing Model

	Precision	Recall	F1-Score
Training	0.965	0.847	0.902
Testing	0.940	0.809	0.869

Evaluasi kinerja per dimensi EUCS menunjukkan variasi performa, dimana untuk dimensi *Ease of Use* menunjukkan performa terbaik dengan *F1-score* uji sebesar 0,930 karena kemampuannya menangkap pola ulasan kemudahan penggunaan secara akurat. Dimensi *Content* juga menunjukkan hasil yang baik dengan *F1-score* uji sebesar 0,916, sedangkan dimensi *Accuracy* dan *Timeliness* menunjukkan performa yang cukup konsisten. Sementara itu, dimensi *Format* menunjukkan performa terendah dengan *F1-score* uji 0,761 akibat variasi bahasa yang lebih luas pada ulasan terkait tampilan antarmuka.

Kestabilan dan konsistensi performa model kemudian divalidasi menggunakan *k-fold cross validation*, yang hasilnya menunjukkan nilai metrik stabil di setiap *fold* dengan rata-rata *F1-score* tertinggi pada dimensi *Ease of Use* sebesar 0,927 dan terendah pada *Format* yaitu 0,766. Keseluruhan hasil evaluasi ini, sebagaimana dirangkum pada Tabel 7, menegaskan bahwa model *Logistic Regression* dengan pendekatan *One-vs-Rest* merupakan model optimal untuk prediksi multilabel dimensi EUCS pada ulasan pengguna aplikasi Neobank.

Tabel 7. Hasil Rata-Rata Cross Validation

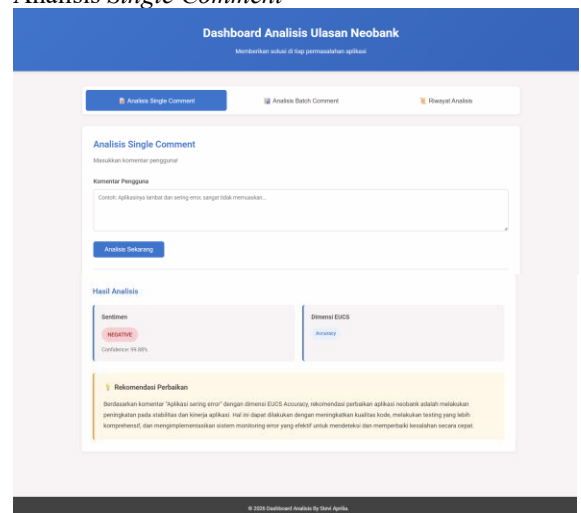
Dimensi EUCS	Precision	Recall	F1-Score
Accuracy	0.953	0.811	0.876
Content	0.986	0.843	0.908
Ease of Use	0.995	0.867	0.927
Format	0.802	0.734	0.766
Timeliness	0.944	0.783	0.856

Penerapan Model Pada Sistem Berbasis Website

Penerapan model optimal pada sistem berbasis web bertujuan untuk mengimplementasikan model analisis sentimen dan prediksi *multilabel* dimensi EUCS ke dalam suatu aplikasi yang mampu memproses komentar pengguna secara otomatis. Implementasi sistem dilakukan melalui arsitektur *client-server* yang mengintegrasikan *framework* Flask pada *backend* dengan menggunakan Python dan antarmuka *frontend* berbasis HTML, CSS, serta JavaScript. Sistem ini mengotomatisasi alur analisis ulasan mulai dari *preprocessing* teks, klasifikasi sentimen menggunakan IndoRoBERTa, prediksi dimensi EUCS secara *multilabel* menggunakan *Logistic Regression*, hingga memberikan rekomendasi perbaikan yang relevan dan kontekstual melalui integrasi LLM dengan menggunakan model Groq/Compound. Simulasi pada lingkungan *localhost* ini membuktikan bahwa model yang dikembangkan telah terintegrasi secara fungsional dan operasional untuk mendukung pengambilan keputusan berbasis data.

Untuk mendukung proses analisis ulasan secara interaktif, Dashboard Analisis Ulasan Neobank dirancang dengan menekankan pada aspek keterbacaan hasil dan kemudahan navigasi. Antarmuka aplikasi terdiri dari beberapa fitur utama sebagai berikut:

a. Analisis Single Comment

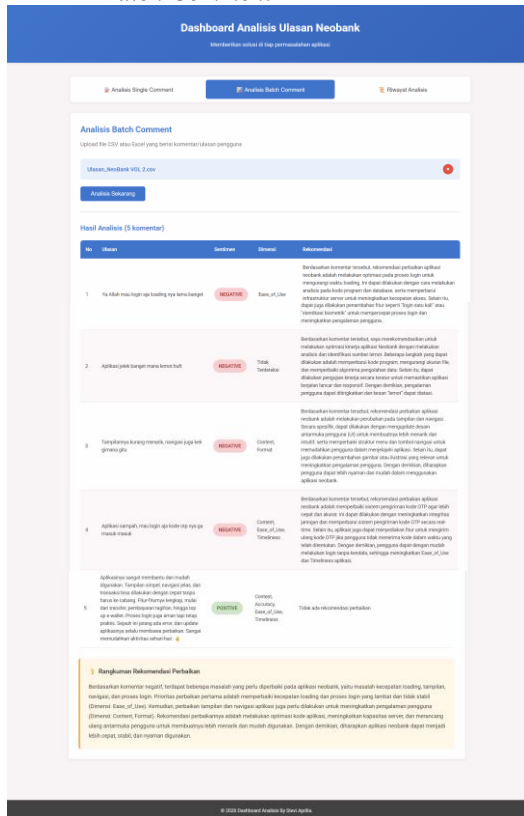


Gambar 4. Halaman Analisis Single Comment

Halaman *Analisis Single Comment* dirancang untuk memfasilitasi evaluasi ulasan secara individual melalui kolom input teks yang memungkinkan pengguna memasukkan komentar secara langsung, kemudian memproses data tersebut melalui tombol *Analisa Sekarang* yang mengirimkan permintaan ke *backend* untuk diproses oleh sistem. Setelah pemrosesan selesai, halaman ini menyajikan hasil analisis secara komprehensif yang mencakup kategori sentimen, nilai *confidence score*, serta prediksi dimensi EUCS, dan dilengkapi dengan bagian *Rekomendasi Perbaikan* yang secara otomatis menyusun narasi solusi berbasis *Large Language Model* (LLM) untuk membantu pengembang memahami langkah mitigasi yang diperlukan.



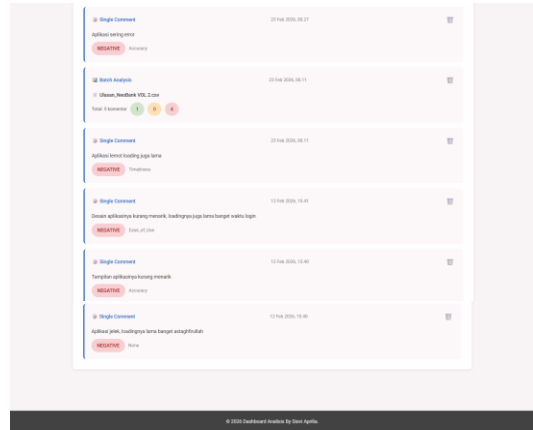
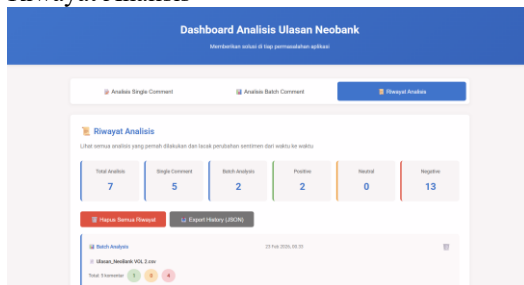
b. Analisis Batch Comment



Gambar 5. Halaman Analisis Batch Comment

Halaman *Analisis Batch Comment* dirancang khusus untuk memproses ulasan dalam jumlah besar melalui fitur unggah berkas berformat CSV atau Excel, yang kemudian dianalisis secara kolektif setelah pengguna menekan tombol *Analisis Sekarang*. Hasil analisis dari setiap komentar, meliputi kategori sentimen, dimensi EUCS, dan rekomendasi perbaikan, disajikan secara sistematis dalam sebuah tabel untuk memudahkan peninjauan. Selain itu, halaman ini dilengkapi dengan bagian *Rangkuman Rekomendasi* yang secara otomatis menyusun kesimpulan umum berdasarkan pola permasalahan yang teridentifikasi dari seluruh kumpulan ulasan tersebut, sehingga memberikan gambaran prioritas perbaikan aplikasi secara lebih efisien.

c. Riwayat Analisis



Gambar 6. Halaman Riwayat Analisis

Halaman Riwayat Analisis berfungsi sebagai pusat penyimpanan data historis untuk memudahkan pengguna melacak hasil analisis yang telah dilakukan sebelumnya tanpa perlu melakukan unggah ulang. Halaman ini dilengkapi dengan fitur *Statistik Analisis* yang menyajikan ringkasan distribusi sentimen dan total aktivitas analisis secara informatif. Selain itu, terdapat *Daftar Riwayat* untuk meninjau kembali rekaman hasil analisis, serta *Fitur Manajemen Riwayat* yang memberikan kendali kepada pengguna untuk menghapus riwayat atau mengeksport data analisis guna keperluan pelacakan dan audit.

Hasil Testing Sistem

Berdasarkan hasil pengujian sistem menggunakan metode *black-box* mengonfirmasi bahwa seluruh fungsi utama Dashboard Analisis Ulasan NeoBank berjalan sesuai perancangan. Fitur utama seperti *input single comment*, *batch upload*, serta proses analisis sentimen dan dimensi EUCS mampu berjalan stabil tanpa *error*. Evaluasi pengguna terhadap fungsionalitas fitur, kemudahan antarmuka, dan persistensi data menunjukkan respons yang positif dengan dominasi penilaian pada skala maksimal.

Dalam tahap komparasi lima arsitektur LLM, model Groq/Compound terbukti paling unggul dalam menghasilkan narasi rekomendasi yang kontekstual dan adaptif terhadap bahasa Indonesia. Berbeda dengan model lain seperti *Owen*, *Orpheus*, atau model *guardrail* (*Llama* dan *GPT-OSS*) yang memiliki kendala pada kualitas bahasa atau kemampuan generatif, *Groq/Compound* memberikan hasil paling solutif. Meski demikian, sistem masih memiliki keterbatasan, terutama terkait ketergantungan pada koneksi API, ketepatan pemetaan dimensi EUCS pada ulasan kompleks, serta ruang lingkup pengujian yang saat ini masih terbatas pada *localhost*.

KESIMPULAN

Penelitian ini menyimpulkan bahwa persepsi kepuasan pengguna aplikasi NeoBank belum didominasi oleh sentimen positif, dengan masih tingginya sentimen negatif yang dipengaruhi signifikan oleh permasalahan kualitas sistem, khususnya pada dimensi *Accuracy* sebagai aspek yang paling berpengaruh. Model klasifikasi *multilabel* menggunakan *Logistic Regression* dengan



pendekatan *One-vs-Rest* dan TF-IDF terbukti optimal dengan performa yang stabil, dibuktikan melalui *F1-score* data uji sebesar 0,869 dan hasil *k-fold cross validation* yang konsisten di setiap *fold*. Integrasi model ini ke dalam sistem berbasis *website* telah berhasil dilakukan dan divalidasi melalui pengujian *black-box* yang menunjukkan fungsionalitas berjalan dengan baik. Selain itu, model *Groq/Compound* ditetapkan sebagai arsitektur *Large Language Model* (LLM) terbaik dalam sistem karena mampu menghasilkan rekomendasi perbaikan yang paling kontekstual, solutif, dan adaptif terhadap variasi ulasan pengguna.

DAFTAR PUSTAKA

- Aini, Fitratul, Fitriani Muttakin, Tengku Khairil Ahsyar, and Eki Saputra. 2023. "Analisis Kepuasan Pengguna Aplikasi DANA Menggunakan." 06(01):65–76.
- Alfarisi, M. Sechan, M. Fathir Bagas, Salwa Aprilia Santi, Ramanda Setiawan, Bakti Ihlusal, and Dwiki Syahrur. 2026. "P ENGEMBANGAN A SISTEN B ELANJA P INTAR B ERBASIS I NTERAKSI S UARA." 62–75.
- Alvin, Alvin, Robet Robet, and Astuti Tarigan. 2025. "Implementasi Chatbot Otomatis Akademik Berbasis Web Menggunakan LLM Dan Rule-Based System Studi Kasus : STMIK Time Implementation of Web-Based Academic Automated Chatbot Using LLM and Rule-Based System Case Study : STMIK Time." (3):651–65. doi: 10.26798/jiko.v9i3.2209.
- Anindya, Fazha Safha, and Yampi R. Kaesmetan. 2025. "Implementasi Metode Bert Dan Svm Pada Analisis Sentimen Game Genshin Impact." *Jurnal Manajemen Informatika Jayakarta* 5(1):52. doi: 10.52362/jmijayakarta.v5i1.1781.
- Djaelani, Abdul Kadir, and Husni Thamrin. 2025. "A Comparative Analysis of Three Transformer Models for Twitter Sentiment: A Case Study of the 'Fufufafa' Account." 1–6. doi: 10.1109/siml65326.2025.11080838.
- Doll, William. J., and Gholamreza. Torkzadeh. 1988. "The Measurement of End-User Computing Satisfaction End-User Satisfaction The Measurement of End-User Computing Satisfaction Professor of MIS and Strategic Management The University of Toledo Gholamreza Torkzadeh Assistant Professor of Information Systems." *MIS Quarterly* 12(2):259–74.
- Ernawati, Ida Ayu, Asif Faruqi, and Virdha Rahma Aulia. 2025. "User Satisfaction Analysis of Financial Technology Applications Using the End User Computing Satisfaction Model." *Journal Binary Digital - Technology* 8(1):757–66. doi: 10.32877/bt.
- Ibadurrahman, Muhammad, Arrasyid Supriyanto, Abdullah Arkananta, Rasendrya Hasan, and Dio Dharmasa. 2025. "Integrasi Mobile Aplikasi Untuk Klasifikasi Harga Laptop Menggunakan Metode Support Vector Classification Dan Logistic Regression JURNAL MEDIA INFORMATIKA [JUMIN]." 6(4):2342–50. doi: <https://doi.org/10.55338/jumin.v6i4.6576>.
- Idris I, Mustofa Y, and Salihi I. 2023. "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)." *Jambura Journal of Electrical and Electronics Engineering* 5:32–35.
- Johri, Amar. 2023. "Exploring Customer Awareness towards Their Cyber Security in the Kingdom of Saudi Arabia : A Study in the Era of Banking Digital Transformation." 2023. doi: 10.1155/2023/2103442.
- Kennedy, Eric B., Eric A. Jensen, and Aaron M. Jensen. 2022. "Methodological Considerations for Survey-Based Research During Emergencies and Public Health Crises: Improving the Quality of Evidence and Communication." *Frontiers in Communication* 6(February):1–7. doi: 10.3389/fcomm.2021.736195.
- Kusumastuti, Rajnaparamitha, Dewi Oktafiani, and Tommy Dwi Putra. 2025. "Implementasi Groq AI Untuk Otomatisasi Feedback Pada Website Evaluasi Kinerja Dosen." 6(1):152–61. doi: 10.35957/algorithm.v5i3.13458.
- Prabawanti, Ratih, and Denny Jean Cross Sihombing. 2023. "Analysis of Factors Affecting User Satisfaction of E-Commerce Applications Using End-User Computing Satisfaction (EUCS) Method." *Journal of Information Systems and Informatics* 5(1):324–32. doi: 10.51519/journalisi.v5i1.437.
- Pratama, Ahmad R., Rio Rizky Aryanto, Arif Taufiq M. Pratama, Universitas Islam Indonesia, Kabupaten Sleman, and Penulis Korespondensi. 2022. "MODEL KLASIFIKASI CALON MAHASISWA BARU UNTUK SISTEM LEARNING-BASED COLLEGE MAJOR RECOMMENDATION SYSTEM." 9(4). doi: 10.25126/jtiik.202294311.
- Ramma, Rafif Rafeda, and I. Kadek Dwi Nuryana. 2025. "Analysis Of User Satisfaction Towards The ShopeePay Application Using The EUCS (End User Computing Satisfaction) Method." 6(2):155–71.
- Rubenstein, Eric, and Sarah Furnier. 2021. "#Bias: The Opportunities and Challenges of Surveys That Recruit and Collect Data of Autistic Adults Online." *Autism in Adulthood* 3(2):120–28. doi: 10.1089/aut.2020.0031.
- Sihombing, Yogie Oktavianus, Nita Vibriyanti Situmorang, Badan Kepegawaian Negara, and Jl Mayjen Sutoyo. 2024. "Prediksi Sentimen Pada Teks Media Sosial Corporate University Menggunakan RoBERTa." *Jl. Mayjen Sutoyo No* 12(12):302–16.
- Sonjaya, Chepy Bagustian, Anis Fitri, Nur Masruriyah, and Dwi Sulistya. 2022. "The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease." 5(2):166–75. doi: <https://doi.org/10.32627/internal.v5i2.595>.
- Sukmana, Febrian Humaidi, Lalu Moh, Nazar Fajri, and Sri Maryanti. 2024. "Mengungkap Persepsi Masyarakat Tentang Pengalaman Layanan Kesehatan Di Rumah Sakit Pemerintah Menggunakan Ulasan Online." 21(2):180–92. doi: <https://doi.org/10.53512/valid.v21i2.401>.
- Supriana, I. Wayan, Kadek Belvanatha, Gargita Satwikandana, and Putu Yuki. 2025. "A Case-Based Reasoning with Artificial Intelligence Approach for Laptop Repair Assistant System." 14(2):349–54.



- Syatriani, Rozzi Kesuma Dinata, and ArRazi. 2025. "Perbandingan Metode Logistic Regression Dan Random Forest Dalam Klasifikasi Penyakit Kulit Multikelas." 10(2):1369–79. doi: <https://doi.org/10.36341/rabit.v10i2.6551>.
- Umar, Savitri, Ahmad Fitriansyah, and Harun Ar-Rasyid. 2023. "Evaluasi Tingkat Kepuasan Pengguna Aplikasi Mobile Banking Livin Bank Mandiri Pada Mahasiswa ITB Swadharma Menggunakan End User Computing Satisfaction (EUCS)." *Jurnal Teknologi Informatika Dan Komputer* 9(2):992–1008. doi: 10.37012/jtik.v9i2.1728.